The Effect of Financial Incentives on Instructional Resource Allocation: A Framed Field Experiment

Austin S. Jennings

University of Delaware

Author Note

Austin S. Jennings, College of Education and Human Development and Lerner College of Business and Economics, University of Delaware

Correspondence concerning this article should be addressed to Austin S. Jennings, 102 Purnell Hall, University of Delaware, Newark, DE 19716

E-mail: austinj@udel.edu

Abstract

Within a principal-agent framework, educational policymakers have the greatest access to the financial incentives necessary to provide extrinsic motivation for teachers' compliance with educational reforms. Contemporary, longitudinal research on financial incentives indicates methods of bonus compensation do not significantly affect student achievement in the United States. However, such studies are unable to directly capture the effect of financial incentives on teachers' existing instructional practices. I develop a framed field experiment to explore pre-service teachers' response to financial incentives and the effect thereof on instructional resource allocation. Aligned therewith, I use a three-level hierarchical generalized linear model to predict the probability a pre-service teacher allocates at least one instructional resource to a given student. Findings suggest pre-service teachers behave strategically based on financial incentive structures, allocating instructional resources to students for whom the expected compensation is greatest. These findings indicate the design of educational policy may affect the distribution of instructional resources and, as a consequence, academic opportunities for students with diverse learning profiles.

Keywords: financial incentives, framed field experiment, instructional resources, educational policy

JEL Classification Codes: A20, C91, C93

The Effect of Financial Incentives on Instructional Resource Allocation:

A Framed Field Experiment

The American public school system may be characterized as a complex network of agency relationships with varied perspectives, goals, and levels of accountability (Ferris, 1992; Labaree, 2010; Levačić, 2009). Agency relationships exist between national and state policymakers, educational leaders, district and school administrators, curricular resource developers, educational resource organizations, stakeholders, parents, and teachers. Within this framework, teachers are ultimately responsible for synthesizing personal and principal desires with pedagogical practices to affect student learning outcomes. Agency problems arise when an agent acts on behalf of one or more principals with imperfect information and uncertainty (Ross, 1973). Given teacher autonomy, the manner in which educational policy is enacted in the classroom is subject to agency problems (Labaree, 2010).

Educational reform efforts vary by principal actor, including: standards and methods of assessment (Taylor, 1994); teacher certification (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005); professional development (Borko, 2004), curriculum design (Ball & Cohen, 1996); and parent involvement (Mattingly, Prislin, McKenzie, Rodriguez, & Kayzar, 2002). Within a principal-agent framework, principal actors may address agency problems through financial schemes designed to incentivized desired outcomes (Sappington, 1991). While any principal actor may initiate educational reform efforts, national and state policymakers have the greatest access to the financial incentives necessary to provide extrinsic motivation for teachers' compliance (Elmore, Ablemann, & Fuhrman, 1996; Labaree, 2010). However, critics contend imperfect information regarding teachers' autonomous actions creates an environment in which financial incentives may result in unanticipated or undesired instructional practices (Murnane & Cohen, 1986).

Current research on financial incentives in the United States typically focuses on student learning outcomes as well as teacher value-added measures and self-reported instructional practices. However, such studies are unable to directly capture the effect of financial incentives on teachers' existing instructional practices. In the current study, I use a framed field experiment to shift the research focus away from measures of student learning outcomes to investigate the effect of financial incentives on student learning opportunities via teachers' allocation of limited instructional resources. This study extends existing literature by providing insight into the manner in teachers' modify existing instructional practices in response to financial incentives.

Literature Review

Merit pay, a form of financial incentive through which compensation is related to specific, measureable outcomes, was first implemented in the United States in 1908 (Protsik, 1995). Merit pay gained formal support from educational leaders in the United States as early as 1916 with subsequent implementation in American public schools including peaks in the 1920s and 1960s (Johnson, 1984). A resurgence of national, political interest in merit pay schemes occurred in the 1980s with the support of President Ronald Reagan and persists with the support of President Barack Obama and Secretary of Education Arne Duncan (Gratz, 2009; Meckler, 2009; Protsik, 1995). In 2010, the United States Department of Education provided funding for financial incentives in 27 states through the Teacher Incentive Fund and evaluated states' structure for implementing performance-based compensation within the context of "Race to the

Top" applications (Fryer, 2013).[1] Aligned with such reform efforts, recent research has focused on the effectiveness of financial incentives in affecting student learning outcomes as well as teachers' attitudes, perceptions, and self-reported instructional practices (Springer & Gardner, 2010).

The Achievement Challenge Pilot Program (ACPP) provided bonus compensation up to $11,200 for teachers in five Little Rock, Arkansas elementary schools between 2004 and 2007 (Winters, Ritter, Greene, & Marsh, 2009); additional financial incentives were provided to school administrators and personnel. Teachers earned bonus compensation based on average student growth as measured by the Iowa Basic of Basic Skills composite score. The authors found the ACCP was associated with a 0.16 standard deviation increase in math achievement and a 0.15 standard deviation increase in reading achievement in a subset of three schools. However, the authors caution schools self-selected into the ACPP and systematically differed from non-ACPP schools. The authors used a 32-item survey instrument to measure teachers' attitudes, perceptions, and behaviors but found no statistically significant differences in self-reported data between ACPP and demographically comparable schools.

The Strategic Compensation Initiative (REACH) provided bonus compensation up to $14,795 for teachers in Austin, Texas schools between 2007 and 2015(Balch & Springer, 2015); additional financial incentives were provided to school principals. Teachers earned bonus compensation based on three categories: student performance, professional development, and retention. The authors found REACH was associated with a 0.18 standard deviation increase in math achievement and 0.06 standard deviation increase in reading achievement as measured by

[1] Although the current study focuses on research and educational policy in the United States, financial incentives have been designed and implemented internationally in countries including: Australia, Chile, Denmark, Finland, Hungary, India, Israel, Mexico, New Zealand, Norway, Portugal, Sweden, Turkey, and the United Kingdom (Fryer, 2013; Woessmann, 2011).

the Texas Essential Knowledge and Skills assessment during the first year of implementation. However, the authors found increases in math and reading achievement did not persist in the second year of implementation and student learning objective-based performance was not significantly associated with teacher value-added estimates. Furthermore, Schmitt, Lamb, Cornetto, and Courtemanche (2014) found student learning objective-based performance was not significantly associated with self-reported use of data to inform instruction.

The Project on Incentives in Teaching (POINT) provided bonus compensation up to $15,000 for fifth through eighth grade mathematics teachers in Nashville, Tennessee schools between 2006 and 2009 (Springer et al., 2012). Teachers earned bonus compensation based on average student growth as measured by the mathematics Tennessee Comprehensive Assessment Program exam. The authors found POINT was not significantly associated with overall or year-specific differences in math achievement. The authors used a series of surveys conducted over three years to create measures of teachers' attitudes, instructional practices, professional development, and perceptions of school environment. The 25-item instructional practices measure included two items related to the effect of financial incentives on teachers' instructional practices.[2] However, the two items did not identify specific practices and were not directly comparable to control group teachers' hypothetical responses.

The Schoolwide Performance Bonus Program (SPBP) provided bonus compensation up to $3,000 per United Federation of Teachers-affiliated teacher in New York City schools between 2007 and 2010 (Marsh et al., 2012); each school independently determined its incentive structure within SPBP guidelines. Schools earned per-teacher bonus compensation based on a

[2] Teachers responded to the two survey items with a number 1 through 4 where 1 represents *strongly disagree* and 4 represents *strongly agree*. Item 1: "I was already working as efficiently as I could before the implementation of POINT, so the experiment is not affecting my work." Item 2: "I have altered my instructional practices as a result of the POINT experiment."

progress report card comprised of three categories: environment, progress, and performance. Fryer (2013) and Marsh et al. (2012) found SPBP was not significantly associated with level-specific student achievement and may decrease achievement in elementary and middle schools. The authors used a series of surveys conducted over three years to create measures of teachers' instructional practices, improvement efforts, collaboration, collegiality, professional development, and instructional leadership. The 15-item frequency of instructional practices measure included nine items related to teachers' use of data to inform instruction.[3]

Marsh et al. (2012) found no statistically significant differences in self-reported frequency of data-informed instructional practices between SPBP and control teachers. Although teachers may not alter the frequency of data-informed instructional practices, such survey items do not capture the manner in which teachers alter existing data-informed instructional practices in response to financial incentive structures. Within accountability frameworks, students are classified based on demographic subgroup, learning characteristics, and proficiency relative to achievement criterion (Booher-Jennings, 2006). Shared student categorizations serve as the foundation for teacher conversations termed *classificatory talk* (Little, 2012). Horn (2007) and Little (2012) contend such discourse provides insight into the manner in which student categorization frames teachers' data interpretation and instructional responses. Furthermore, Booher-Jennings (2005) and Neal and Schanzenbach (2010) find teachers engage in *educational triage* practices through which students' distance from classificatory thresholds are significantly associated with access to instructional resources and learning opportunities.

[3] Teachers responded to the nine survey items with a number 1 through 5 where 1 represents *never* and 5 represents *almost daily*. Items 1-9: "I use test score data to – (1) identify individual students who need additional assistance; (2) set learning goals for individual students; (3) tailor instruction to individual students' needs; (4) develop recommendations for tutoring or other educational support services; (5) assign or reassign students to groups within my class; (6) identify topics requiring more or less emphasis in instruction; (7) encourage parent involvement in student learning; (8) identify areas where I need to strengthen my content knowledge or skills; (9) reflect on and discuss teaching and learning with my inquiry team or other teachers, coaches, etc."

The current study expands on existing education research by addressing the relationship between financial incentives, data-informed instruction, and teachers' instructional resources allocation. I use a framed field experiment to explore (1) how financial incentives affect teachers' instructional resource allocation and (2) how this relationship is mediated by student achievement data. This study represents the first known application of experimental economics methodology to gather insight into teachers' behavioral response to financial incentives and student achievement data. The framework developed in this study may be used to explore principal-agent relationships in education, teachers' interpretation and implementation of educational policy, as well as implications for student learning opportunities and outcomes.

Experimental Design

Framed field experiments engage a purposefully selected subject pool in an experiment with field context in task or information set (Harrison & List, 2004). With respect to subject pool, the current experiment sampled pre-service teachers to control for participant age, academic coursework, method of certification, and teaching experience. The sample consists of 43 pre-service teachers enrolled in an elementary education program at a large, Mid-Atlantic research university. The sample was 98% female and 0% minority with a mean age of 20.3 years of age. The sample was recruited from courses designated for junior- and senior-year undergraduates to ensure all participants were formally enrolled in the elementary education program. 91% of participants were enrolled in their junior year; 23% and 77% of participants had a concentration in mathematics and special education, respectively. Participants engaged in classroom observations, developed standards-based lesson plans, and co-taught individual lessons through coursework prior to participation in the current study. The sample represents over two-thirds of the pre-service teacher cohort enrolled in the identified courses at the time of the experiment.

The framed field experiment was designed as an iPad application using Xcode version 7.1.1. Participants engaged in a field-contextualized task in which they were responsible for allocating limited instructional resources among students. In each round of the experiment, participants allocated 20 units of instruction to a class of ten students. Each round represents a school year as participants were given a new set of ten students each round while the outcome of previous rounds could inform current decisions. Units of instruction represent instructional resources or learning opportunities in addition to general instruction such as tutoring, pull-out instruction, remediation or enrichment instruction, and extension activities (Booher-Jennings, 2006). Simulated classrooms consisted of ten students to ensure competition for limited instructional resources while limiting the demands on participants' information processing.

Participants were provided field-contextualized student information based on the Smarter Balanced Assessment Consortium (SBAC; 2014) accountability framework. All aspects of the general experimental interface reflect disaggregated data included in the SBAC report provided to teachers in the state in which the experiment was conducted.[4] Participants were provided each student's name, predicted score range on a standardized assessment, and assessment-based classification. Student names were randomly generated at the beginning of each round and mapped to randomized student achievement data in alphabetical order. Students' initial understanding was depicted on a continuum ranging from 2400 to 2800 points with an increase

[4] Only first names were included in the experimental interface to limit the amount of information participants must process. The continuum of understanding and predicted score range reflect the SBAC report in color and design. The overall score range was determined as a composite of Mathematics and English Language Arts ranges across upper elementary and middle school grade levels. Classifications of understanding are drawn from SBAC achievement level descriptors. Growth in classification, as presented to participants with financial incentives, is not explicitly included in SBAC reports but was included in the experimental interface to reinforce the calculation of bonus compensation.

in points representing a greater depth of understanding of instructional content. Each student's predicted score range spanned 20 points on the continuum from which a final score was drawn based on a uniform distribution. The continuum was divided into four equal parts representing the four SBAC classifications of understanding: minimal, partial, adequate, and thorough.

Participants were randomly assigned to a salary or financial incentive condition upon initiating the iPad application. The value of participants' decisions were expressed in experimental dollars (E$) to reflect values consistent with contemporary salary schedules and financial incentives research. Participants earned E$40,000 (converts to $0.50) per round under the salary condition. This value reflects the average salary for teachers with a bachelor's degree and zero years of teaching experience in the state in which the experiment was conducted. Under the financial incentive condition, participants earned a maximum bonus of E$15,500 (converts to $0.65) per round based on a two-part incentive scheme. These values reflect the maximum annual bonus compensation teachers could earn during the REACH and POINT incentive programs. Participants earned E$1,000 per increase in student classification and an additional E$1,500 per student with at least an adequate final understanding.[5] Instructional videos for both conditions were embedded in the iPad application (see Appendix A for verbatim transcripts). Participants' comprehension of instructions was reinforced by three to four interactive questions and participants could review the instructional video at any point throughout the experiment.

The experiment consisted of four parts, each comprised of six rounds. In each round, participants allocated 20 units of instruction to their students. Each unit of instruction increased student achievement by 10 points. Participants could devote one unit to an individual student to increase his or her achievement by 10 points. Alternately, participants could devote one unit to

[5] See Appendix A for two illustrations of bonus compensation calculation in the verbatim transcript of instructions for participants assigned to the financial incentives condition.

whole class instruction to increase all ten students' achievement by 1 point each. To this end, allocating twenty units of instruction to whole class instruction is equivalent in outcome to devoting two units of instruction to each of the ten students. At the end of each round, each student's final understanding is determined by a score drawn from a uniform distribution based on his or her predicted score range.

Part order was randomized upon initiating the iPad application; however, parts will be discussed in order of complexity throughout the subsequent analyses. In Part 1, initial scores were centered 10 points below the threshold for partial, adequate, and thorough understanding. Participants could ensure all students demonstrate classification-based growth by devoting two units of instruction to each student, one unit to each student and ten units to whole class instruction, or all 20 units to whole class instruction. Alternately, participants could support differentiated growth through an alternate allocation of instructional resources. Part 1 was developed to measure participants' allocation preferences and was incorporated in the current study.

In Part 2, initial scores were centered 50 points below classification thresholds at the midpoint of the minimal, partial, and adequate ranges of understanding. Part 2 was developed to reflect teachers' potential to differentiate student understanding between, but not within, classifications. Part 2 deviates from SBAC reporting conventions as the accountability framework provides data that could be used to differentiate within categorizations including individualized student reports. However, research indicates teachers classificatory talk generally refers to students' membership in broad categorizations which mask differences within categorizations (Horn, 2007). To this end, Part 2 measures teachers' instructional resource allocation provided limited student achievement data.

In Parts 3 and 4, initial scores were randomly assigned within the minimal, partial, and adequate ranges of understanding. Initial scores were centered at 10-point increments such that participants could ensure classification-based growth with a discrete quantity of units of instruction if desired. Parts 3 and 4 were developed to reflect teachers' potential to differentiate student understanding both between and within classifications. To this end, Parts 3 and 4 more accurately reflect SBAC reporting conventions and measure teachers' instructional resource allocation provided detailed student achievement data.

Empirical Model

Hierarchical linear modeling is employed in education research to extend beyond traditional, level-specific questions and explore structural relationships between variables (Alkharusi, 2011; Raudenbush, 1988). I employed a three-level hierarchical generalized linear model (HGLM) predicting the probability a pre-service teacher allocates at least one unit of instruction to a given student. A student received the equivalent of one unit of instruction if his or her teacher (1) allocated him or her at least one student specific unit of instruction or (2) allocated the class at least ten units of whole class instruction. The three-level model is comprised of students nested within classrooms nested within teachers.

Level-1 models the relationship between the probability of instructional resource allocation and student-level variables. The level-1 structural model assumes the mathematical form:

$$\eta_{ijk} = \pi_{0jk} + \pi_{1jk}(threshold_{ijk}) + \pi_{2jk}(partial_{ijk}) + \pi_{3jk}(adequate_{ijk}) + \pi_{4jk}(threshold*partial_{ijk}) + \pi_{5jk}(threshold*adequate_{ijk})$$

where:

η_{ijk} = the log odds for allocation of at least one unit of instruction for student *i*, in classroom *j*, with teacher *k*

π_{0jk} = the model intercept

π_{pjk} = the regression coefficient for variable *p*

Threshold distance (*threshold*) measures the difference between students' initial predicted achievement score and the minimum final score needed to demonstrate one level of classification-based growth with certainty. Threshold distance is measured in 10-point increments to represent the units of instruction required to ensure a given student increases in classification. Classification of understanding (*partial* and *adequate*) indicates students' initial level of academic achievement. Minimal understanding served as the omitted category and no student began with a thorough understanding in any round of the experiment. Interaction terms (*threshold**partial* and *threshold**adequate*) measure the differential effect of threshold distance based on students' initial classification. All level-1 variables are group-mean centered.

Level-2 models the relationship between the probability of instructional resource allocation and classroom-level variables. The level-2 structural model assumes the mathematical form:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(detailed_{jk}) + r_{0jk}$$

$$\pi_{pjk} = \beta_{p0k} \qquad \text{for } p = 1, 4, 5$$

$$\pi_{pjk} = \beta_{p0k} + \beta_{p1k}(detailed_{jk}) \qquad \text{for } p = 2, 3$$

where:

β_{00k} = the classroom intercept for the model intercept

β_{01k} = the regression coefficient for detailed information

β_{p0k} = the classroom intercept for variable *p*

β_{p1k} = the regression coefficient for the interaction between detailed information and variable *p*

r_{0jk} = the random effect for classroom *j* with teacher *k*

Detailed information (*detailed*) represents the accuracy of student achievement data. As described within the experimental design, participants were provided limited student achievement data in Part 2 and detailed student achievement data in Parts 3 and 4. In Part 2, participants could differentiate student understanding between, but not within, classifications. In Parts 3 and 4, participants could differentiate student understanding both between and within classifications. All level-2 variables are group-mean centered.

Level-3 models the relationship between the probability of instructional resource allocation and teacher-level variables. The level-3 structural model assumes the mathematical form:

$\beta_{00k} = \gamma_{000} + \gamma_{001}(incentive_k) + u_{00k}$

$\beta_{01k} = \gamma_{010}$

$\beta_{p0k} = \gamma_{p00} + \gamma_{p01}(incentive_k)$ for p = 1, 2, 3

$\beta_{p1k} = \gamma_{p10}$ for p = 2, 3

$\beta_{p0k} = \gamma_{p00}$ for p = 4, 5

where:

γ_{000} = the teacher intercept for the model intercept

γ_{001} = the regression coefficient for financial incentives

γ_{p10} = the teacher intercept for detailed information

γ_{p00} = the teacher intercept for variable *p*

γ_{p01} = the regression coefficient for the interaction between financial incentives and variable *p*

u_{00k} = the random effect for teacher k

Financial incentives (*incentive*) represents whether the participant was randomly assigned to the incentive-based, as opposed to constant salary, condition. All level-3 variables are grand-mean centered.

Results

Table 1 presents distributional statistics and univariate correlations between the predictors and criterion. Based on standards established in Hinkle, Wiersma, and Jurs (2003), bivariate associations reveal pre-service teachers' instructional resource allocation is negatively correlated with threshold distance suggesting an increase in a students' distance from the next classification is statistically associated with decreased access to instruction. Pre-service teachers' instructional resource allocation is positively correlated with minimal understanding and negatively correlated with partial and adequate understandings suggesting minimal understanding is statistically associated with increased access to instruction. Pre-service teachers' instructional resource allocation is positively correlated with detailed information suggesting more accurate achievement data is statistically associated with a more equitable distribution of instruction. Finally, pre-service teachers' instructional resource allocation is negatively correlated with incentives suggesting financial incentives are statistically associated with a less equitable distribution of instruction.

Table 2 presents estimates of fixed and random unit-specific effects from the multilevel model predicting the probability pre-service teachers allocate at least one unit of instruction to a given student. Multilevel modeling provides a framework for exploring the structural relationship between covariates while reducing cluster bias in parameter estimates and correcting standard errors for the non-independence of structural data (Guo & Zhao, 2000). The model is based on a sample of 7740 students, nested within 774 classrooms, nested within 43 teachers. I estimated the multilevel model using HLM7 with heteroskedasticity-robust standard errors and an unstructured covariance matrix.

The main effect for classroom- and teacher-level variables indicates both detailed achievement data and financial incentives have statistically significant associations with the probability of instructional resource allocation. Pre-service teachers are 83% more likely to allocate at least one unit of instruction to the average student when provided detailed achievement data. This finding suggests pre-service teachers are more likely to equitably distribute instructional resources when they are able to differentiate between and within achievement classifications. However, pre-service teachers are 47% less likely to allocate at least one unit of instruction to the average student when provided financial incentives. This finding suggests pre-service teachers are more likely concentrate instructional resources on individual students when bonus compensation is earned based on students' classification-based growth and final understanding. Exploration of student-level variables indicates additional, interaction effects for classroom- and teacher-level variables.

Pre-service teachers are 11% less likely to allocate at least one unit of instruction to the average student given a one unit increase in threshold distance. Furthermore, a one unit increase in threshold distance is associated with a 29% reduction in likelihood if that student has a partial or adequate, as opposed to minimal, initial understanding. These findings suggest pre-service teachers are more likely to support students close to classification-based growth with a smaller effect for students with a minimal initial understanding. Pre-service teachers earning financial incentives are 42% less likely than those earning a constant salary to allocate at least one unit of instruction to the average student given a one unit increase in threshold distance. This relationship increases to a 54% reduction in likelihood for students with a partial or adequate, as opposed to minimal, initial understanding. These findings suggest pre-service teachers earning bonus compensation, as opposed to a constant salary, are more responsive to students' threshold distance.

Pre-service teachers are 57% less likely to allocate at least one unit of instruction to the average student with a partial, as opposed to minimal, initial understanding. This relationship increases to a 79% reduction in likelihood for students with an adequate initial understanding. These findings suggest pre-service teachers are most likely to allocate instructional resources to students with a minimal understanding and least likely to allocate instructional resources to students with an adequate understanding. This relationship persists after controlling for the effect of detailed achievement data. Pre-service teachers with detailed achievement data are 62% and 80% less likely to allocate at least one unit of instruction to the average student with partial and adequate initial understanding, respectively. However, this relationship reverses after controlling for the effect of financial incentives. Pre-service teachers earning financial incentives are 4.8 times more likely to allocate at least one unit of instruction to the average student with partial initial understanding. These findings suggest financial incentives significantly affect pre-service teachers' instructional resource allocation among students with differential initial achievement.

Discussion

This study explores the relationship between financial incentives and teachers' instructional practices. Contemporary, longitudinal research on financial incentives indicates methods of bonus compensation do not significantly affect student achievement in the United

States (Fryer, 2013; Marsh et al., 2012; Springer et al., 2012). Furthermore, current research indicates financial incentives do not significantly affect teachers' self-reported instructional practices including the frequency of data-informed instruction (Marsh et al., 2012). However, the current use of survey instruments does not capture the manner in which educational policy affects teachers' existing instructional practices. Through a framed field experiment, I have examined pre-service teachers' data-informed instructional practices and the effect of financial incentives on teachers' instructional resource allocation.

Pre-service teachers are more likely to allocate instructional resources to students with minimal initial understanding and those closest to classification-based growth. Furthermore, the negative effect of threshold distance is greater for students with partial or adequate initial understanding. These findings are consistent with existing literature on achievement-based classification, instructional practices, and student growth. Brown and Saks (1987) found teachers employ compensating strategies by focusing resource allocation on those students with the lowest initial achievement. Although pre-service teachers were significantly more likely to allocate instructional resources to students with minimal initial understanding, this result is slightly mediated by threshold distance. In the current study, threshold distance has a larger relative effect for students with partial and adequate initial understanding. This behavior is consistent with Booher-Jennings (2005) and Neal and Schanzenbach (2010) who found teachers engage in educational triage, providing increased access to instructional resources and learning opportunities to students closest to classification-based growth.

Further analysis of the relationship between achievement data and instructional practices indicates pre-service teachers are more likely to allocate instructional resources to the average student when provided data that differentiated student understanding both within and between achievement classifications. However, classification effects persisted despite the provision of detailed information. This finding further supports the presence of educational triage and compensating strategies, suggesting pre-service teachers allocated fewer instructional resources to students closest to the threshold for partial understanding and allocated excess resources to those students with the lowest initial understanding.

The two-part incentive scheme had two effects consistent with Murnane and Cohen's (1986) contention that financial incentives will affect instructional practices and may result in unanticipated or undesired outcomes. First, financial incentives increased the effect of threshold distance on the likelihood of instructional resource allocation. To this end, the introduction of bonus compensation increased educational triage by incentivizing instructional resource allocation to students closest to classification-based growth. This finding illustrates the distributional effect of financial incentives on instructional resource allocation. Although each unit of instruction resulted in a 10 point increase in student achievement regardless of initial understanding, pre-service teachers were less likely to allocate resources to those students furthest from classification-based growth. Alternate incentive schemes may result in similar, potentially undesired instructional practices. For example, the POINT incentive scheme provided bonus compensation based on students' growth relative to peers with similar initial understanding (Springer et al., 2012). This bonus compensation structure may incentivize the resource allocation to students with the greatest perceived return to investment.

Second, financial incentives altered the effect of initial classification on the likelihood of instructional resource allocation. To this end, the introduction of bonus compensation redirected pre-service teachers' compensating strategies toward a focus on students with partial understanding. This finding further illustrates the distributional effect of financial incentives on instructional resource allocation. To this end, salaried pre-service teachers' instructional practices narrowed the range of final understanding by focusing resource allocation on the lowest performing students. Alternately, incentivized pre-service teachers maximized the number of students with an adequate final understanding, creating a divide between the lowest performing students and their peers. This finding is consistent with those of Neal and Schanzenbach (2010), who found teachers subject to proficiency-based accountability systems devote additional resources to students near proficiency criterion.

Policy Implications

Within a principal-agent framework, Murnane and Cohen (1986) contend teacher autonomy creates an environment in which financial incentives may result in unanticipated or undesired instructional practices. The current study demonstrates the use of experimental economics methodology and framed field experiments to explore teachers' interpretation and implementation of educational policy. Findings suggest pre-service teachers behave strategically based on financial incentive structures, allocating instructional resources to students for whom the expected compensation is greatest. Although the intent of financial incentive schemes such as that employed in the current study may be to motivate teachers and promote student proficiency, the structure thereof has implications for the distribution of student learning opportunities and outcomes.

Differences in resource allocation between salaried and incentivized pre-service teachers highlight the relationship between teachers' instructional practices and interpretation of success as defined by educational policies. Whereas salaried pre-service teachers employed compensating strategies by allocating resources to students with the lowest initial understanding, incentivized pre-service teachers engaged in educational triage by allocating resources to students closest to the proficiency threshold. These findings indicate the design of educational policy may affect the distribution of instructional resources and, as a consequence, academic opportunities for students with diverse learning profiles. To this end, policymakers should extend educational policy analyses beyond broad student achievement and value-added measures to include analyses of the manner in which teachers alter existing instructional practices in accordance with policy interpretations. Experimental economics methodology and framed field experiments provide an efficient context for educational policy design and evaluation through exploration of teachers' interpretation and implementation.

Limitations and Future Research

This study represents the first known application of experimental economics methodology to gather insight into teachers' behavioral response to financial incentives and student achievement data. Findings are limited in external validity by the sample, financial incentive structure, and aspects of experimental design. While the sample is representative of the students enrolled in the program from which the sample was drawn, it is not likely to be representative of all new teachers in the state. The sample was 98% female and 0% minority with all pre-service teachers' concentration in either mathematics or special education. Future research may expand on the current study by drawing a larger, more representative sample of teachers. The financial incentive structure was designed to reward teachers for students' classification-based growth and proficiency in final understanding. Given the finding teachers align resource allocation with their interpretation of success as defined by educational policy, future research may expand on the current study by exploring alternate financial incentive designs. Finally, aspects of the study were designed to limit the amount of information pre-service teachers were required to process in the context of the experiment. For example, classrooms were comprised of only ten students, participants had a fixed number of units of instruction, units of instruction had a constant return, and all instruction occurred in a single period of time. Future research may expand on the experimental design by increasing the number of students per class to more accurately reflect the complexity of environment in which teachers act. Participants could be given the opportunity to increase their units of instruction at a cost to explore teachers' willingness to devote time outside the normal instructional day in response to financial incentives. The return per unit of instruction could be altered based on students' initial understanding to explore the effect of differential learning rates. Alternately, the return per unit of instruction could decrease with each additional unit devoted to a given student to explore the effect of diminishing marginal returns. Despite limitations, this study develops a framework within a principal-agent perspective that has the potential to explore teachers' interpretation and implementation of educational policy.

References

Alkharusi, H. (2011). Hierarchical linear models: Applications in educational assessment research. *Educational Research Journal*, *26*(1), 41-69.

Balch, R., & Springer, M. G. (2015). Performance pay, test scores, and student learning objectives. *Economics of Education Review*, *44*, 114-125.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6-8+14.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231-268.

Booher-Jennings, J. (2006). Rationing education in an era of accountability. *The Phi Delta Kappan*, *87*(10), 756-761.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3-15.

Brown, B. W., & Saks, D. H. (1987). The microeconomics of the allocation of teachers' time and student learning. *Economics of Education Review*, *6*(4), 319-332.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), 1-51.

Elmore, R. F., Ablemann, C. H., & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98).Washington, D.C.: The Brookings Institution.

Ferris, J. M. (1992). School-based decision making: A principal-agent perspective. *Educational Evaluation and Policy Analysis*, *14*(4), 333-346.

Fryer, R. G., Jr. (2013). Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, *31*(2), 327-371.

Gratz, D. B. (2009). Special topic: The problem with performance pay. *Multiple Measures*, *67*(3), 76-79.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, *26*, 441-462.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009-1055.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin

Horn, I. S. (2007). Fast kids, slow kids, lazy kids: Framing the mismatch problem in mathematics teachers' conversations. *The Journal of the Learning Sciences*, *16*(1), 27-79.

Johnson, S. M. (1984). Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review*, *54*(2), 175-185.

Labaree, D. F. (2010). *Someone has to fail: The zero-sum game of public schooling*. Cambridge, MA: Harvard University Press.

Levačić, R. (2009). Teacher incentives and performance: An application of principal-agent theory. *Oxford Development Studies*, *37*(1), 33-46.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*(2), 143-166.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., & Peng, A. (2012). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Santa Monica, CA: RAND Corporation.

Mattingly, D. J., Prislin, R., McKenzie, T. L., Rodriguez, J. L., & Kayzar, B. (2002). Evaluating evaluations: The case of parent involvement programs. *Review of Educational Research*, *72*(4), 549-576.

Meckler, L. (2009, March 11). Education push includes merit pay. The Wall Street Journal. Retrieved from http://www.wsj.com/articles/SB123668036405881929

Murnane R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review*, *56*(1), 1-17.

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, *92*(2), 263-283.

Protsik, J. (1995). *History of incentive pay and incentive reform*. Washington, DC: Educational Resources Information Center.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: *A review. Journal of Educational Statistics*, *13*(2), 85-116.

Ross, S. A. (1973). Theory of agency: The principal's problem. *The American Economic Review*, *63*(2), 134-139.

Sappington, D. E. M. (1991). Incentives in principal-agent relationships. *The Journal of Economic Perspectives, 5*(2), 45-66.

Schmitt, L. N. T., Lamb, L. M., Cornetto, K. M., & Courtemanche, M. (2014). *AISD REACH program update, 2012-2013: Student learning objectives*. Austin, TX: Austin Independent School District.

Smarter Balanced Assessment Consortium. (2014). Smarter Balanced Assessment Consortium reporting system: Reporting system overview. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/06/Reporting-System-Overview.pdf

Springer, M. G., & Gardner, C. D. (2010). Teacher pay for performance: Context, status, and direction. *The Phi Delta Kappan*, *91*(8), 8-15.

Springer, M. G., Ballou, D., Hamilton, L., Le, V.-N., Lockwood, J. R., McCaffrey, D. F., Pepper, M., Stecher, B. M. (2012). *Final report: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal*, *31*(2), 231-262.

Winters, M. A., Ritter, G. W., Greene, J. P., & Marsh, R. (2009). Student outcomes and teacher productivity and perceptions in Arkansas. In M. G. Springer (Ed.), *Performance incentives: Their growing impact on American K-12 education* (pp. 273-293). Washington, DC: Brookings Institution Press.

Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review*, 30, 404-418.

Table 1

Distributional Statistics and Bivariate Correlations for Criterion and Predictors

Variable	*N*	*M*	*SD*	1	2	3	4	5	6
allocation	7740	0.532	.499	-.212***	.309***	-.084***	-.224***	.098***	-.065***
Level-1									
1. *threshold*	7740	5.900	2.324		.034**	.044***	-.081***	-.030**	.000
2. *minimal*	7740	0.317	.465			-.537***	-.446***	.025*	.000
3. *partial*	7740	0.383	.486				-.516***	-.024*	.000
4. *adequate*	7740	0.300	.458					.000	.000
Level-2									
5. *detailed*	774	0.667	.471						.000
Level-3									
6. *incentive*	43	0.512	.500						

Note: * p < .05, ** p < .01, *** p < .001; *allocation* = 1 if a student receives the equivalent of at least one unit of support, = 0 otherwise.

Table 2

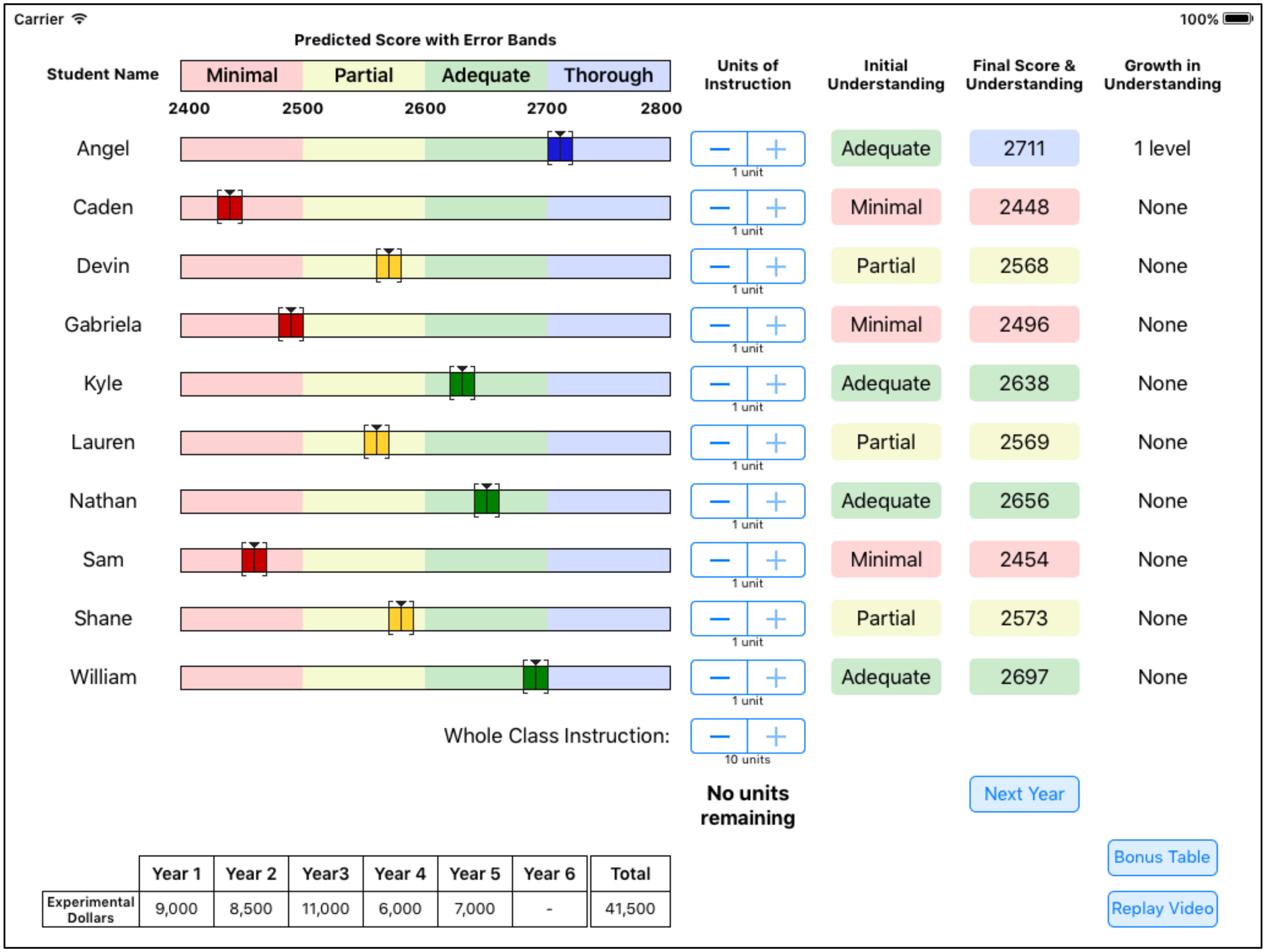
Unit-Specific Multilevel Model Predicting the Probability of Instructional Resource Allocation

MRCM fixed effect	Variable	Estimate(robust *SE*)	Odds Ratio	*p*-value
Intercept	γ_{000}	0.349(.119)**	1.148	.005
Level-1 effects				
threshold	γ_{100}	-0.120(.045)**	0.887	.007
partial	γ_{200}	-0.856(.351)*	0.425	.015
adequate	γ_{300}	-1.573(.346)***	0.207	< .001
*threshold*partial*	γ_{400}	-0.217(.051)***	0.805	< .001
*threshold*adequate*	γ_{500}	-0.226(.059)***	0.797	< .001
Level 2 effect				
detailed	γ_{010}	0.605(.125)***	1.831	< .001
Level-2 cross level interactions				
detailed x *partial*	γ_{210}	-0.721(.336)*	0.486	.032
detailed x *adequate*	γ_{310}	-0.652(.315)*	0.521	.039
Level 3 effect				
incentive	γ_{001}	-0.626(.237)*	0.535	.012
Level-3 cross-level interactions				
incentive x *threshold*	γ_{101}	-0.430(.075)***	0.650	< .001
incentive x *partial*	γ_{201}	3.240(.628)***	25.537	< .001
incentive x *adequate*	γ_{301}	2.680(.586)***	14.580	< .001
MRCM random effect	Variable	Variance Component	$\chi^2(df)$	*p*-value
variance of classroom effects	r_0	.747***	1615.206(730)	< .001
variance of teacher effects	u_{00}	.780***	334.196(41)	< .001

Note: * $p < .05$, ** p < .01, *** p < .001; MCRM = multilevel random coefficient model.

Figure 1

Experimental Interface



Appendix A

Verbatim Transcript of Instructions: Constant Salary

Welcome to an experiment in the economics of decision making. During the experiment, you will act as the teacher of ten students. You will decide how to spend 20 units of instruction on your ten students. Each unit of instruction will improve your students' understanding as measured by a standardized assessment. You will earn a salary by spending all 20 units of instruction. Your salary will be expressed in experimental dollars. At the end of each part of the experiment, your experimental dollars will be converted into US dollars. When you complete the experiment, you will receive and keep the US dollars you earned. Let's take a look at how you will make decisions and earn money.

This is the screen you will use to make your decisions. Here, you can see your students' names. These bars show your students' understanding. The highlighted area shows each student's predicted score range on a standardized assessment. At the end of each year, your students' score will be randomly selected from the colored portion of his or her predicted score range. Here, you can see your students' initial understanding. You will use these buttons to add and subtract units of individualized instruction. Each unit of instruction is worth ten performance points. You can give all ten performance points to a single student by adding a unit with the student-specific button. Alternately, you can give one performance point to each of your ten students by adding a unit with the whole class instruction button. As you add and subtract units, your students' predicted score range will change. You can see how many hours you have given each student under his or her button. You can also see how many hours you have left at the bottom of the column. If you would like to erase your decisions, click Reset Year. After you decide how to spend all 20 units, click Submit.

When you click Submit, you will see each student's final score and a color indicating final understanding. For example, April began with a partial understanding.
April's final score is 2765. She has a thorough understanding so the background color is blue. We knew April would have a thorough understanding because her entire possible score range was in the blue region of her bar. After each year, the experimental dollars you earn will be recorded in the table at the bottom of the screen. After you review the result of your decisions, click Next Year.

You will get a new group of ten students with new initial scores each year. Once you have completed all six years, click Continue to move to the next part of the experiment. The experiment has four parts. At the end each part, you will see the total amount of money you earned in US dollars. If you would like to watch all or part of this video again, click Replay Video at any time.

If you have any questions, please raise your hand. Otherwise, let's get started!

Verbatim Transcript of instructions: Financial Incentives

Welcome to an experiment in the economics of decision making. During the experiment, you will act as the teacher of ten students. You will decide how to spend 20 units of instruction on your ten students. Each unit of instruction will improve your students' understanding as measured by a standardized assessment. Your students' performance will determine the size of your bonus. Your bonus will be expressed in experimental dollars. At the end of each part of the experiment, your experimental dollars will be converted into US dollars. When you complete the experiment, you will receive and keep the US dollars you earned.

This table shows how your bonus will be calculated based on your students' initial and final understanding. You earn 1,000 experimental dollars for each level a student grows and an additional 1,500 experimental dollars if the student's final understanding is at least adequate. For example, imagine a student named April who begins with a minimal understanding. If April ends with a minimal understanding, you will not earn a bonus for her performance because she did not demonstrate growth and was below an adequate understanding. If April ends with a partial understanding, you will earn 1,000 experimental dollars because she demonstrated one level of growth. If April ends with an adequate understanding, you will earn 3,500 experimental dollars: 2,000 because she demonstrated two levels of growth and an additional 1,500 because she had at least an adequate final understanding. If April ends with a thorough understanding, you will earn 4,500 experimental dollars: 3,000 because she demonstrated three levels of growth and an additional 1,500 because she had at least an adequate final understanding.

Now, imagine a student named May who begins with an adequate understanding. If May ends with an adequate understanding, you will earn 1,500 experimental dollars for her performance because she did not demonstrate growth but had at least an adequate final understanding. If May ends with a thorough understanding, you will earn 2,500 experimental dollars: 1,000 because she demonstrated one level of growth and an additional 1,500 because she had at least an adequate final understanding. The performance bonus from all ten of your students will be added together to determine your overall bonus. Let's take a look at how you will make decisions and earn money.

This is the screen you will use to make your decisions. Here, you can see your students' names. These bars show your students' understanding. The highlighted area shows each student's predicted score range on a standardized assessment. At the end of each year, your students' score will be randomly selected from the colored portion of his or her predicted score range. Here, you can see your students' initial understanding. Remember, this is important because the amount of your bonus depends on each student's initial and final understanding. You will use these buttons to add and subtract units of individualized instruction. Each unit of instruction is worth ten performance points. You can give all ten performance points to a single student by adding a unit with the student-specific button. Alternately, you can give one performance point to each of your ten students by adding a unit with the whole class instruction button. As you add and subtract units, your students' predicted score range will change. You can see how many hours you have given each student under his or her button. You can also see how many hours you have left at the bottom of the column. If you would like to erase your decisions, click Reset Year. After you decide how to spend all 20 units, click Submit.

When you click Submit, you will see each student's final score, a color indicating final understanding, and the levels of growth between initial and final understanding. For example, April began with a partial understanding. April's final score is 2759. She has a thorough understanding so the background color is blue. We knew April would have a thorough understanding because her entire possible score range was in the blue region of her bar. Thorough is two levels higher than partial, so April has grown by two levels. If you would like to review the value of your decisions, click Bonus Table at any time. After each year, the experimental dollars you earn will be recorded in the table at the bottom of the screen. After you review the result of your decisions, click Next Year.

You will get a new group of ten students with new initial scores each year. Once you have completed all six years, click Continue to move to the next part of the experiment. The experiment has four parts. At the end each part, you will see the total amount of money you earned in US dollars. If you would like to watch all or part of this video again, click Replay Video at any time.

If you have any questions, please raise your hand. Otherwise, let’s get started!